



# Crowd Flow Characterization with Optimal Control Theory

Pierre Allain, Nicolas Courty, Thomas Corpetti

## ► To cite this version:

Pierre Allain, Nicolas Courty, Thomas Corpetti. Crowd Flow Characterization with Optimal Control Theory. Ninth Asian Conference on Computer Vision (ACCV 2009), Jul 2009, Xi'an, China. pp.279–290. hal-00493443

**HAL Id: hal-00493443**

**<https://hal.science/hal-00493443>**

Submitted on 22 Jun 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Crowd Flow Characterization with Optimal Control Theory

Pierre Allain<sup>1</sup>, Nicolas Courty<sup>1</sup>, and Thomas Corpetti<sup>2</sup>

<sup>1</sup> Université Européenne de Bretagne – VALORIA / UBS

{Pierre.Allain,Nicolas.Courty}@univ-ubs.fr

<sup>2</sup> CNRS / COSTEL – INRIA Rennes Thomas.Corpetti@irisa.fr

**Abstract.** Analyzing the crowd dynamics from video sequences is an open challenge in computer vision. Under a high crowd density assumption, we characterize the dynamics of the crowd flow by two related information: velocity and a disturbance potential which accounts for several elements likely to disturb the flow (the density of pedestrians, their interactions with the flow and the environment). The aim of this paper is to simultaneously estimate from a sequence of crowded images those two quantities. While the velocity of the flow can be observed directly from the images with traditional techniques, this disturbance potential is far more trickier to estimate. We propose here to couple, through optimal control theory, a dynamical crowd evolution model with observations from the image sequence in order to estimate at the same time those two quantities from a video sequence. For this purpose, we derive a new and original continuum formulation of the crowd dynamics which appears to be well adapted to dense crowd video sequences. We demonstrate the efficiency of our approach on both synthetic and real crowd videos.

## 1 Introduction

Analyzing crowd video sequences has recently revealed to open specific and original problems in computer vision. Direct applications consider the design of safety systems for public confined or opened spaces. In this case, the goal of a surveillance system is to be able to give an information of the flow of persons at a given time in a given situation. From this information, one can infer useful statistics about dangerous areas such as bottlenecks or narrow passages. Automatic surveillance system can also trigger alarms whenever abnormal or dangerous situations are detected. It is also noticeable that such tools participate to our comprehension of crowd phenomena. As an example, Helbing and colleagues [1] have recently build a new theory on crowd dynamics based on an analysis of video recordings of the annual pilgrimage in Makkah. Such a study of the crowd behavior from real data enables to hold out critical locations (*i.e.* areas with high density and pressure) of a scene and can in addition contributes to the elaboration of accurate simulation models. Another original recent application is related to computer graphics and the production of digital effects: Courty and Corpetti [2] designed a data-driven crowd animation system based on the velocity fields acquired with optical flow techniques from an input crowd video sequence.

The coupling of crowd dynamics and real data exhibits very promising results and has opened a rich area of research. This paper is a contribution in that direction. We argue that the apparent motion information is intrinsically insufficient to characterize the dynamics of the flow since the lack of motion in the image can be interpreted as a null density or a large congestion area where people are likely to be injured. We define a substantially complete crowd flow analysis as the extraction from the sequence of *i*) time-consistent motion fields and *ii*) an associated disturbance potential. The motion field is a rich dynamical descriptor of the flow which can be related to the velocity of flow. The disturbance potential accounts for several physical quantities such as the density or the pressure in the flow. This information is crucial to extract sensible and potentially dangerous areas. Although an important number of approaches are available to measure the apparent velocity field from images sequences in various situations, the estimation of the disturbance potential is a critical problem and is still an open domain of research. This component is indeed tricky to observe directly from images. It is nevertheless intuitive that this potential influences the motion field: in a natural way, human beings tend to avoid over-concentrated or high-pressure areas, and their velocities are directly influenced by the surrounding person concentration.

The original contribution of this paper is to use recipes from optimal control theory [3] and variational assimilation [4], originally used in the context of meteorology, to define a new tool for the characterization of the crowd flow. Such techniques enable to estimate a (potentially high dimensional) system state driven with a dynamic model known up to some noise. A key advantage relies on the ability to measure unobserved parameters that control the dynamic model. As such, it is thoroughly adapted to the problem we are dealing with. The definition of a system based on variational assimilation especially requires *i*) a dynamic model related to the motion field and the disturbance potential and *ii*) an observation operator that links our data (images) to some components of system state (motion fields). Among others, we propose in this paper a new and crude physical model for crowd dynamics and apply it to estimate time-consistent informations of image sequences of human crowds.

The remainder of the paper is organized as follows: after the presentation of related work in the context of crowd flow analysis from video, we give an overview of our method (Section 3.1). Section 3.2 presents our modeling of the problem and implementation issues. Before concluding, Section 4 exposes our experimentations on both synthetic (with the associated ground truth) and real crowd sequences.

## 2 Related Work

Analysis of crowd video sequences are generally focused toward two distinct problems: the counting of people in the crowd and the detection of abnormal situations where accidents are likely to occur. The counting issue generally yields the questions of *i*) background subtraction and *ii*) feature tracking. Concerning the tracking, the choice of the features to extract is determinant. Typical methods are based on appearance models [5, 6] that exhibit different sensibilities to inter persons occlusions. The temporal and spatial consistency of the tracked features can be obtained through clustering methods [7]. In [8], Brostow and Cipolla successfully argue that only the apparent motion in image

space is relevant to singularize individual in the crowd flow. When several hundreds of pedestrians are present in the crowd most of the conventional tracking methods (like Kalman filters or particle filtering [9, 10]) fail, because the degradation of the visual features related to single individuals disturbs the analysis. Moreover, the large induced state space yields computationally too expensive problems. In those cases, the analysis of the crowd sequence may amount to the analysis of a crowd **flow** that have global properties and may be treated as a whole. Works related to this class usually tend to solve the different problems of event detection or changes in the flow rate [11, 12]. The analysis usually takes as input the apparent motion in the image space (optical flow). In [12], unsupervised feature clustering is used to define normal motion patterns, and HMMs are used to detect particular situations. The method proposed by Ali and Shah in [11] allows to segment the crowd flow with regions of substantially different dynamics by examining the Lagrangian coherent structures in the flow. In some sense, this Eulerian perception of the crowds dynamics (that assumes that the crowd can have fluid-like properties) opposes to a Lagrangian view of the individual tracking problem.

In our method, we propose to use a physical crowd model to guide the analysis knowledge. This physical model is new and can be related to the continuous formulations of crowd dynamics such as the one of Hughes [13]. In a recent work [14], Ali and Shah also use an *a priori* knowledge on the crowd dynamics by using a scene structured based force model that guides a tracking process. This work differs from ours in the sense that our goal is not the individual tracking of people in the scene, but rather a characterization of the entire crowd flow.

Given a dynamic model related to a phenomenon, different filtering processes have been developed to extract time-consistent parameters. Recently, Papadakis *et al.* have exploited the variational assimilation principle to successfully extract time-consistent and high-dimensional state spaces (dense motion fields, curves, physical parameters such as vorticity, ...) directly from images [15, 16]. The variational assimilation methods perform in batch and allow to explicitly enforce a (more or less confident) dynamic model to the variables to recover. Their framework is expressed by means of an *adjoint formulation*: adjoint variables are introduced and enable to compute easily the gradient of the cost-function. The resulting algorithm consists of iterating a forward integration of the evolution model and a backward integration of the adjoint evolution model guided by a discrepancy measurement between the state variable and the available noisy observations. This efficient procedure authorizes to refine an initial condition (which can be low confident) as well as the deviations *wrt.* the dynamic model.

In this paper, we suggest to define a dynamic model for crowds that couples the velocity to the disturbance potential. Recalling that the main difficulty of crowd analysis concerns the estimation of this last potential, this model will be a support to estimate this quantity using variational assimilation. This is the scope of the next section.

### 3 Estimation by coupling observations/model with variational assimilation

#### 3.1 Overview of the method

We recall here that our objective is to estimate at the same time the apparent motion of the crowd and its disturbance potential in the image sequence. An overall schema is given in Figure 1. We take as input the original images and two user-defined information: the eventual position of obstacles and some predefined destination areas in the image. These two information are combined to compute a potential function that conveys information on the optimal directions of displacements for the crowd. From the input images are also derived some initializations for our algorithm as well as the observations (that mainly consist in the apparent motion between image pairs). These are used in the assimilation process, that tries to match, through an iterative process, the observations and the evolution of the dynamical process. As a result, a complete sequence of velocity and disturbance potential are computed.

We present some background on variational assimilation in the next section (3.2), while our model, along with implementation issues, is thoroughly described in the ending part of this section (3.3).

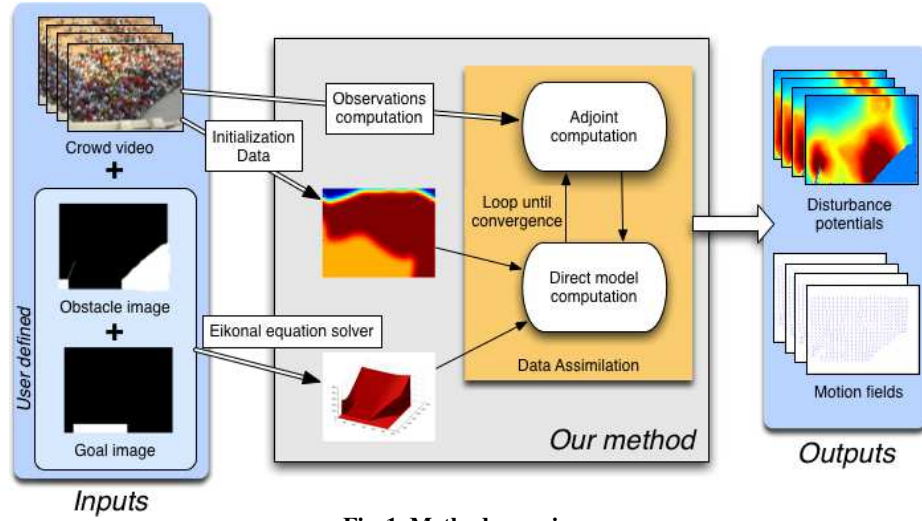


Fig. 1. Method overview.

#### 3.2 Variational assimilation

In this section, the key points required for the comprehension of the variational assimilation are introduced. A complete and detailed presentation would be out of the scope of this paper. For details, we refer the reader to [3, 4].

Our problem consists in recovering, from an initial condition, a system's state  $\mathcal{X}$  partially observed and driven with an approximately known dynamic. This formalizes as finding  $\mathcal{X}(x, t)$ , for any location  $x$  at time  $t \in [t_0, t_f]$ , that satisfies the system:

$$\frac{\partial \mathcal{X}}{\partial t}(x, t) + \mathbb{M}(\mathcal{X}(x, t)) = \epsilon_m(x), \quad (1)$$

$$\mathcal{X}(x, t_0) = \mathcal{X}_0(x) + \epsilon_n(x), \quad (2)$$

$$\mathcal{Y}(x, t) = \mathbb{H}(\mathcal{X}(x, t)) + \epsilon_o(x, t), \quad (3)$$

where  $\mathbb{M}$  is the non-linear operator relative to the dynamics,  $\mathcal{X}_0$  is the initial vector at time  $t_0$  and  $(\epsilon_n, \epsilon_m)$  are (unknown) variables relative to noise on the dynamics and the initial condition respectively. Besides, noisy measurements  $\mathcal{Y}$  of the unknown state are available through the non-linear operator  $\mathbb{H}$  up to  $\epsilon_o$ . To estimate the system's state a common methodology consists in defining a cost-function  $\mathcal{J}$  based on the three previous relations to minimize. The evaluation of  $\mathcal{X}$  can be done by canceling the gradient of this cost function. Unfortunately, the estimation of such gradient is in practice unfeasible for large system's state since it requires to compute perturbations along all the components of  $\mathcal{X}$ . A way to cope this difficulty, firstly proposed by Lions in [3], is to write an *adjoint formulation* of the problem. It can be shown that this yields the following algorithm:

1. Starting from  $\tilde{\mathcal{X}}(x, t_0) = \mathcal{X}_0(x)$ , perform a *forward* integration:  $\frac{\partial \tilde{\mathcal{X}}}{\partial t} + \mathbb{M}(\tilde{\mathcal{X}}) = 0$
2.  $\tilde{\mathcal{X}}$  being available, **find the adjoint variables**  $\lambda(x, t)$  with the *backward* equation:

$$\lambda(t_f) = 0 ; \quad - \frac{\partial \lambda}{\partial t}(t) + \left( \frac{\partial \mathbb{M}}{\partial \mathcal{X}} \right)^\dagger \lambda(t) = \left( \frac{\partial \mathbb{H}}{\partial \mathcal{X}} \right)^\dagger R^{-1}(\mathcal{Y} - \mathbb{H}(\tilde{\mathcal{X}}))(t) \quad (4)$$

3. **Update the initial condition** :  $d\mathcal{X}(t_0) = B\lambda(t_0) + d\mathcal{X}(t_0)$ ;
4.  $\lambda$  being available, **find the state space**  $d\mathcal{X}(t)$  from  $d\mathcal{X}(t_0)$  with the *forward* integration

$$\frac{\partial d\mathcal{X}}{\partial t}(t) + \left( \frac{\partial \mathbb{M}}{\partial \mathcal{X}} \right) d\mathcal{X}(t) = Q\lambda(t) \quad (5)$$

5. **Update** :  $\tilde{\mathcal{X}} = \tilde{\mathcal{X}} + d\mathcal{X}$
6. **Loop** to step 2 until convergence

where the matrices  $B, Q, R$  are relative to the covariance of the errors  $(\epsilon_m, \epsilon_n, \epsilon_o)$ ,  $\left( \frac{\partial \mathbb{M}}{\partial \mathcal{X}} \right)$  and  $\left( \frac{\partial \mathbb{H}}{\partial \mathcal{X}} \right)$  are the *linear tangent operators* of  $\mathbb{M}$  and  $\mathbb{H}$  respectively<sup>3</sup> and  $\left( \frac{\partial \mathbb{M}}{\partial \mathcal{X}} \right)^\dagger$  and  $\left( \frac{\partial \mathbb{H}}{\partial \mathcal{X}} \right)^\dagger$  their *adjoint operators*<sup>4</sup>. Intuitively, the adjoints variables  $\lambda$  contain information about the discrepancy between the observations and the dynamic model. They are computed from a current solution  $\tilde{\mathcal{X}}$  with the backward integration (4) that implicates both observations and dynamical operators. This deviation information between data/model is then used to refine the initial condition (step 3) and to recover the system state through an imperfect dynamic model where errors are  $Q\lambda$  (step 4). Note that if the dynamic is supposed to be perfect (like in many physical applications), the associated covariance  $Q$  is null and the algorithm only refines the initial condition.

<sup>3</sup> The linear tangent of an operator  $\mathbb{A}$  is the Gâteaux derivative :  $\lim_{\beta \rightarrow 0} \frac{\mathbb{A}(X + \beta\theta) - \mathbb{A}(X)}{\beta}$

<sup>4</sup> The adjoint  $\mathbb{A}^\dagger$  of a linear operator  $\mathbb{A}$  on a space  $\mathcal{D}$  is such as  $\forall x_1, x_2 \in \mathcal{D}$ ,  
 $\langle \mathbb{A}x_1, x_2 \rangle = \langle x_1, \mathbb{A}^\dagger x_2 \rangle$

From the previous algorithm, a complete assimilation system is then defined with *i*) a dynamic model  $\mathbb{M}$ ; *ii*) an observation operator  $\mathbb{H}$ ; *iii*) an initial condition and *iv*) the error covariance matrices  $B$ ,  $Q$  and  $R$ . The next section defines all these components for our problem.

### 3.3 Dynamic model, observations and covariance

**Proposed dynamic model for crowd behavior** The aim of this part is to design a simple dynamic model for crowds that will be used for the assimilation. The system's state  $\mathcal{X}$  is composed of the two components of interest that are the velocity field  $\mathbf{v} = (u, v)^T$  and of the disturbance potential of the crowd  $D$  ( $\mathcal{X} = (u, v, D)^T = (\mathbf{v}, D)^T$ ). Let us define a model for the velocity evolution.

*Velocity modeling* In order to get a prior knowledge of the displacement of the crowd, we assume that all human share the same goal and that the topology (obstacles) of the analyzed scene is available. In a first place our methodology is thus restricted to image sequences exhibiting one main flow of pedestrians. Reasonably assuming that each pedestrian aims at minimizing their travel time to their objectives, the optimal direction at a given location can be modeled as the gradient of a potential function  $\Phi$  defined over the whole domain  $D$ . This potential is the solution of the classical Eikonal equation which has among others been widely used in the context of path planning [17, 18]. For a given scene, we then derive an optimal field  $\mathbf{V} = (U, V)^T = \nabla\Phi$  of the pedestrians that corresponds to the theoretical normalized direction of a pedestrian without any constraint. If now the pedestrians evolve in a crowded environment, we assume that if their velocity differs from the optimal direction, this is due to a disturbance into the scene (density, pressure, ...). Therefore, we propose the following dynamical model:

$$\mathbf{v}(\mathbf{x}, t) = \alpha \left( \mathbf{V}(\mathbf{x}, t) - \underbrace{\beta \nabla D(\mathbf{x}, t)}_{\text{disturbance repulsion}} \right) \quad (6)$$

where  $\alpha$  and  $\beta$  are two constant coefficients that depend on the global speed of the scene.

*Disturbance potential modeling* As for the disturbance potential modeling, we simply assume that this scalar quantity is transported by the motion field and is also eventually diffused along time. This corresponds to a simple physical equation of transport of a scalar. It then obeys to a classical advection-diffusion relation:

$$\frac{\partial D(\mathbf{x}, t)}{\partial t} + \mathbf{v}(\mathbf{x}, t) \cdot \nabla D(\mathbf{x}, t) = \delta \Delta D(\mathbf{x}, t). \quad (7)$$

where  $\delta$  is a small diffusing parameter. Finally, the complete dynamical system of  $\mathcal{X} = (\mathbf{v}, D)^T$  reads (with  $(\bullet) = (\mathbf{x}, t)$ ):

$$\boxed{\begin{bmatrix} \mathbf{v}(\bullet) \\ \frac{\partial D(\bullet)}{\partial t} \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & \alpha\beta\nabla \\ 0 & \mathbf{v}(\bullet) \cdot \nabla - \beta\Delta \end{bmatrix}}_{\mathbb{M}(\mathcal{X})} \begin{bmatrix} \mathbf{v}(\bullet) \\ D(\bullet) \end{bmatrix} = \begin{bmatrix} \alpha\mathbf{V}(\bullet) \\ 0 \end{bmatrix} + \boldsymbol{\epsilon}_m} \quad (8)$$

To suppress the obstacle influence in the computation of the gradient  $\nabla$ , we have used non-symmetric finite-difference in their neighborhood. Concerning the Laplacian operator  $\Delta$  related to the diffusion in (7), we have applied an anisotropic operator that do not diffuse into the obstacles. This dynamic model  $\mathbb{M}$  is non-linear due to the advection term  $\mathbf{v}(\bullet) \cdot \nabla$  that depends on the density. In practice, at a given iteration  $n$ , the velocity  $\mathbf{v}$  used for the advection is the one obtained at iteration  $n - 1$  so that the operator is linear. The associated tangent linear  $(\frac{\partial \mathbb{M}}{\partial \mathbf{x}})$  is then itself. The analytical expression of the adjoint  $(\frac{\partial \mathbb{M}}{\partial \mathbf{x}})^\dagger$  is more tricky to obtain but in our implementation, we have used the fact that its discrete version is the transpose of the discrete version of  $(\frac{\partial \mathbb{M}}{\partial \mathbf{x}})$  [19].

Let us now turn to the observations of the state variables.

**Observations: velocity based on optical-flow** As mentioned above, only the motion fields  $\mathbf{v}$  can be accurately observed from the images, the disturbance potential being a tedious quantity to estimate. Starting from the well-known optical flow constraint equation (ofce), one can assume, to cope with the aperture problem, that the unknown optic flow vector at a location  $\mathbf{x}$  is constant within some neighborhood of size  $n$  [20]. The motion field respects then:

$$K_n * \underbrace{\left( \frac{\partial I(\mathbf{x}, t)}{\partial t} + \nabla I(\mathbf{x}, t) \cdot \mathbf{v}(\mathbf{x}, t) \right)}_{dI/dt} \approx 0, \quad (9)$$

where  $I$  stands for the luminance function and  $K_n$  is a Gaussian kernel of standard deviation  $n$ . From the previous relation, the observation system  $\mathcal{Y}(\mathbf{x}, t) = \mathbb{H}(\mathbf{x}, t)\mathcal{X}(\mathbf{x}, t) + \epsilon_o$  can be defined with (noting  $I_\bullet = \partial I / \partial \bullet$ ):

$$\mathcal{Y}(\mathbf{x}, t) = K_n * I_t(\mathbf{x}, t) \text{ and } \mathbb{H}(\mathbf{x}, t) = [-K_n * I_x(\mathbf{x}, t), -K_n * I_y(\mathbf{x}, t), 0]. \quad (10)$$

This observation operator involves only the motion field. This means that the correction on the disturbance potential will uniquely be achieved by relying on motion observations. From a computational point of view, this operator is linear. The associated tangent linear and adjoints are then derived in the same way than previously.

**Covariances and initialisations** For the initialization, we only need to get the disturbance potential since the corresponding initial velocity field is obtain from (6). The choice of this density depends on the scene to be analyzed. In our experiments, it was roughly set manually and filtered with a Gaussian kernel. Noting that the assimilation process refines this initialization, this latter can be only issued from a coarse and manual estimation.

The covariance matrix of the initial condition  $B$  and the covariance matrix of the dynamic model parameter  $Q$  have been fixed to constant diagonal matrices (no spatial prior on the validity of the model and the initial density are available). Concerning the observation covariance  $R$ , we have used  $R = R_{max} + (R_{min} - R_{max})(1 - \exp(-\|\nabla I\|/\sigma^2))$ . This states that when the image brightness does not contain gradients, the usual ofce is not valid and the covariance is maximal. At the opposite, when high gradients appears, the ofce is confident and  $R$  is low.



## 4 Experimentations and discussion

In this section we present some experimental results obtained with our method. We first begin by giving some comparison elements in a synthesized case. We then present the results obtained on a real crowd video. This part is concluded by a discussion on the proposed method.

For all the following experimentations, the value used the optimal control system were  $B = 0.5I_d$ ,  $Q = 0.1I_d$ ,  $R_{min} = 0.5$ ,  $R_{max} = 5$ ,  $\sigma = 9$ .

### 4.1 Validation

**Ground truth generation** Our goal here is to compare the obtained results to some ground truth. In order to get a flexible validation pipeline, we used synthesized crowd scenes. Designing our simulation framework was done with the following constraints: *i)* a totally different crowd simulation model than the one used for assimilation *ii)* the video should present realistic details in terms of visual appearance and fine pedestrians motion (such as arms and legs balancing). It is then possible to confront the virtual ground truth to the results of the assimilation process.

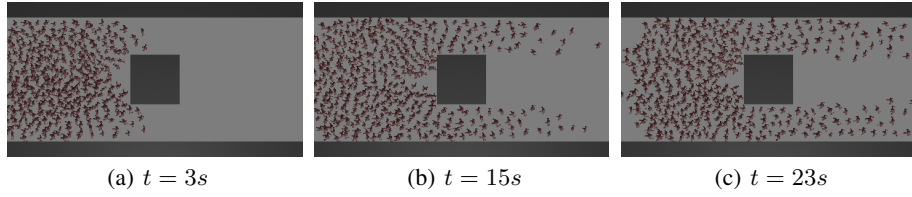
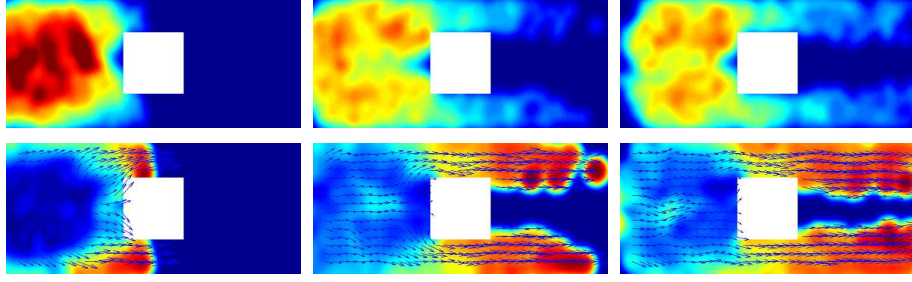


Fig. 2. Synthesized crowd scene

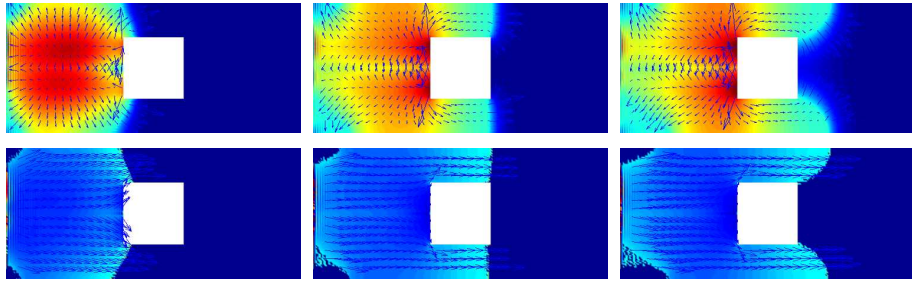
The virtual sequence has been generated using an agent-based crowd simulation model slightly derived from Helbing's model [21], and virtual characters including walking motions acquired through motion capture (see Figure 2). The density maps  $\rho(\mathbf{x}, t)$  and the velocity fields  $\mathbf{U}(\mathbf{x}, t)$  are computed from the agent model to the grid using a Gaussian kernel regularisation (we used  $\sigma = 0.5$ ), and will be considered as the truth.

**Dynamic model results** The integration of the proposed simulated dynamic model provides important information. First, one can see (Figure 4) that it is able, through the disturbance potential, to locate the places where the pedestrians are assumed to be effectively the most disturbed. But in counterpart, velocities tend to quickly decrease over the flow, and do not match to the supposed freedom of move once the obstacle is overtaken.

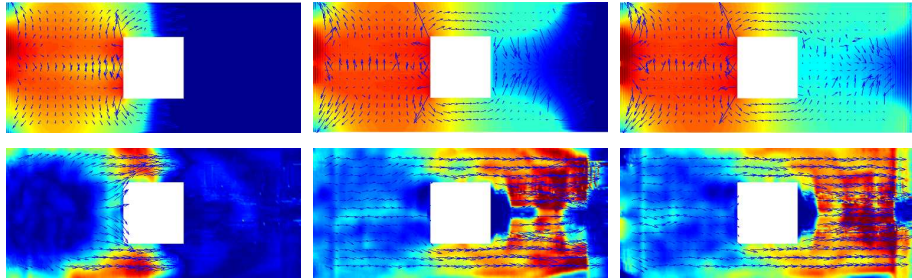
The assimilation, Figure 5, improves the results. As shown on  $D$  maps, the values fit much better the truth density maps, which is an important part of the disturbance potential, in time and space, and particularly after the pedestrians have passed the obstacle.



**Fig. 3.** Ground truth (agent-based model) - Upper line: Density maps (0 - dark blue, to 5 - dark red)  $\text{pedestrians}/\text{m}^2$  - Lower line: Velocity fields and norm (0 to 1  $\text{m}/\text{s}$ )



**Fig. 4.** Simulated dynamic model - Upper line: Disturbance ( $D$ ) maps (0 to 5) with associated gradient (that informs about potential repulsion) - Lower line: Velocity fields and norm

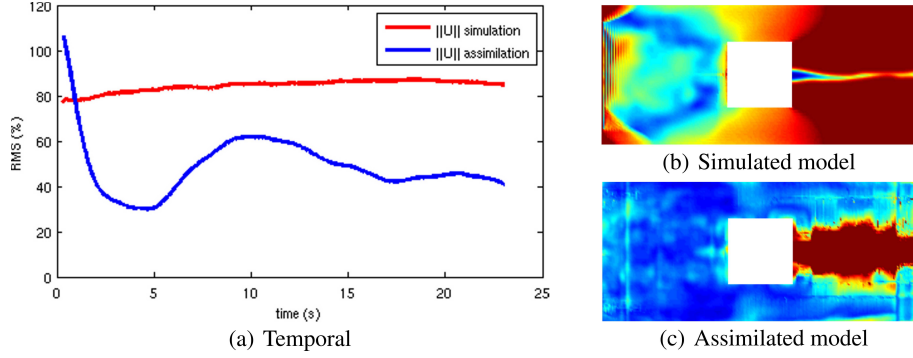


**Fig. 5.** Assimilated dynamic model - Upper line: Disturbance ( $D$ ) maps (0 to 5) with associated gradient - Lower line: Velocity fields and norm (0 to 1)

The velocities norms also fit well to the truth, and the correlation between high disturbance and low speed is clearly effective. Only the back part of the obstacle remains uncorrectly managed.

**RMS comparison** Because of its strong link with pedestrian disturbance, it is relevant to compare the norm of the velocity to the ground truth. In this purpose, we used the standard RMS function. One can see Figure 6, that the assimilation of  $D$  greatly

improved the results as compared to the simple model simulation. The global error is almost lowered by a factor 2. However, the model, as said before, is not well suited to the back part of the obstacle.



**Fig. 6.** Temporal RMS (a) - Spatial RMS (0 to 100%) at  $t = 23s$  on velocity norm (b) and (c)

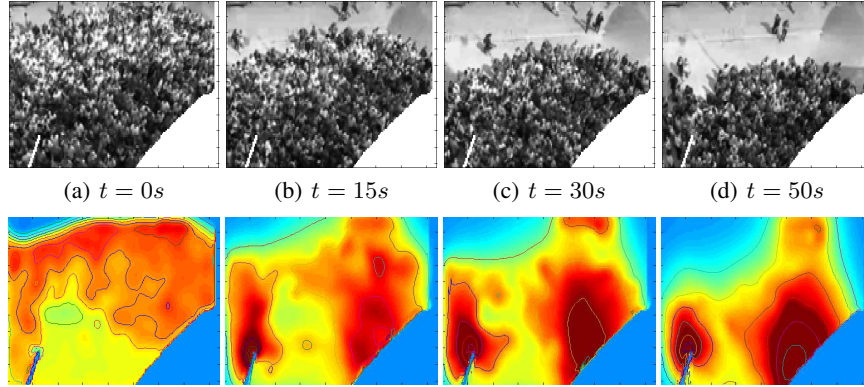
Let us now turn to experiments on real sequences.

#### 4.2 Experimentation on real crowd

The real sequence shows a crowd entering a railway station in the Principality of Monaco (Figure 7). This example is interesting since a variety of phenomena are present: a continuous flow at the beginning followed by a compression of some peoples in the left part of the images. In addition, the limit of the door is a barrier that creates an opposite flux in the crowd flow. In this example, our method has detected two sensible areas where the disturbance potential is growing larger : the end of the barrier and the wall on the right of the image. This is very informative for safety engineers, since it allows to highlight potential risky zones. From an online surveillance system point of view, our method can detect critical disturbance elevations and thus would allow to trigger alarms. It is also possible to connect this information in some motion pattern detector such as presented in [12]. Those aspects have been left as perspectives. Let us remark here that the problem of validation is difficult since no ground truth is available. Nevertheless, from the state-of-the-art on crowd behavior, our estimations seem coherent.

#### 4.3 Discussion

One drawback of our method is that it requires an overhead view of the crowd flow, which is not always available in surveillance system. However, this constraint can be partially alleviated provided that a camera calibration can be given, allowing to project back the observations in a correct frame. Secondly, our method is well adapted to dense crowds, where it is possible to assume that each individual is driven by an underlying flow. With only a few pedestrians in the video, this assumption does not hold anymore.



**Fig. 7.** (a) to (d) Images of the real sequence - (e) to (h) Estimated disturbance potential maps

The range of validity of our system, in terms of crowd density, has thus to be clearly established. Also, we have only treated the case where only one type of crowd (one common goal) is present in the image sequence. We believe that it is possible to handle with our dynamical model more than one flux of people (distinct goals), thus allowing to some extent a segmentation of the crowd flow. Finally, the disturbance potential is a combination of several physical quantities such as density or pressure. We plan to use a more sophisticated physical crowd model to estimate each of these quantities separately in a variational assimilation framework. Those aspects are part of our future works.

## 5 Conclusion and Future Works

We have presented in this paper a complete framework dedicated to the analysis of dense crowd video sequences. Our approach relies on the coupling of observed data extracted from the image sequences and an ad-hoc crowd dynamics model that capture the intrinsic relation between velocity and a disturbance potential (related to density, pressure, ...) in the crowd flow. This allows to derivate an efficient framework that computes from a crowd image sequences both the disturbance potential and the velocity fields over the entire sequence. The estimated disturbance potential has proved to be very interesting in highlighting the main characteristics of a scene. This can serve as inputs for event detection system or participate to our global comprehension on the underlying dynamics of human crowds. Future works will consider a more precise investigation of the disturbance potential to separate from this latter density, pressure and other quantities. We will also investigate sequences containing distinct fluxes of persons.

**Acknowledgements** This work was supported by the french *Région Bretagne AS-Foule* project and by the University of South Brittany *Math-Stic* department.

## References

1. Helbing, D., Johansson, A., Al-Abideen, H.: Dynamics of crowd disasters: An empirical study. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* **75**(4) (2007) 046109
2. Courty, N., Corpetti, T.: Crowd motion capture. *CAVW, proc. of CASA 2007* **18**(4–5) (2007) 361–370
3. Lions, J.: *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag (1971)
4. Le-Dimet, F., Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus* (1986) 97–110
5. Yang, T., Li, S., Pan, Q., JingLi: Real-time multiple object tracking with occlusion handling in dynamic scenes. In: *CVPR, San Diego, USA* (June 2005) 406–413
6. Zhao, T., Nevatia, R.: Tracking multiple humans in crowded environment. In: *CVPR, Washington, DC, USA* (2004) 406–413
7. Rabaud, V., Belongie, S.: Counting crowded moving objects. In: *CVPR, New York* (June 2006) 705–711
8. Brostow, G.J., Cipolla, R.: Unsupervised bayesian detection of independent motion in crowds. In: *CVPR, NYC* (June 2006) 594–601
9. Okuma, K., Taleghani, A., de Freitas, N., Little, J., Lowe, D.: A boosted particle filter: Multitarget detection and tracking. In: *ECCV. Volume 3021 of LNCS., Prague, Czech Republic* (June 2004) 28–39
10. Khan, Z., Balch, T., Dellaert, F.: MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements. *IEEE PAMI.* **28**(12) (December 2006) 1960–1972
11. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: *CVPR, Minneapolis, Minnesota* (June 2007) 1–6
12. Andrade, E., Blunsden, S., Fisher, R.: Modelling crowd scenes for event detection. In: *ICPR, Hong Kong, China* (August 2006) 175–178
13. Hughes, R.L.: The flow of human crowds. *Annual revue of Fluid. Mech.* **20**(10) (2003) 169–182
14. Ali, S., Shah, M.: Floor fields for tracking in high density crowd scenes. In: *ECCV.* (October 2008) 1–14
15. Papadakis, N., Corpetti, T., Mémin, E.: Dynamically consistent optical flow estimation. In: *Proc. Int. Conf. Comp. Vis.(ICCV’07), Rio de Janeiro, Brazil* (October 2007)
16. Papadakis, N., Mémin, E.: A variational technique for time consistent tracking of curves and motion. *Journal of Mathematical Imaging and Vision* (2008) available online first.
17. Polymenakos, L., Bertsekas, D., Tsitsiklis, J.: Implementation of efficient algorithms for globally optimal trajectories. *IEEE Trans. on Automatic Control* **43** (February 1998) 278–282
18. Kimmel, R., Sethian, J.: Optimal algorithm for shape from shading and path planning. *J. of Math. Ima. and Vis.* **14**(3) (2001) 237–244
19. Talagrand, O.: *Variational assimilation. Adjoint equations*. Kluwer Academic Publishers (2002)
20. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *IJCAI., Vancouver, Canada*, (1981) 674–679
21. Helbing, D., Farkas, I., Vicsek, T.: Simulating dynamical features of escape panic. *Nature* **407**(1) (2000) 487–490